

Two-Dimensional Mellin and Mel-Cepstrum for Image Feature Extraction

Serdar ÇAKIR and A. Enis ÇETİN*

Department of Electrical and Electronics Engineering
Bilkent University, 06800, Ankara, Turkey
{cakir,cetin}@bilkent.edu.tr

Abstract. An image feature extraction method based on two-dimensional (2D) Mellin cepstrum is introduced. The concept of one-dimensional (1D) mel-cepstrum which is widely used in speech recognition is extended to two-dimensions both using the ordinary 2D Fourier Transform and the Mellin transform in this article. The resultant feature matrices are applied to two different classifiers (Common Matrix Approach and Support Vector Machine) to test the performance of the mel-cepstrum and Mellin-cepstrum based features. Experimental studies indicate that recognition rates obtained by the 2D mel-cepstrum based method are superior to the recognition rates obtained using 2D PCA and ordinary image matrix based face recognition in both classifiers.

1 Introduction

Mel-cepstral analysis is one of the most popular feature extraction technique in speech processing applications including speech and sound recognition and speaker identification. Two-dimensional (2D) cepstrum is also used in image registration and filtering applications [1, 2]. To the best of our knowledge 2-D mel-cepstrum which is a variant of 2D cepstrum is not used in image feature extraction, classification and recognition problems. The goal of this paper is to define the 2-D mel-cepstrum and 2-D Mellin-cepstrum and show that they are viable image representation tools.

Fourier-Mellin transform (FMT) is a mathematical feature extraction tool which is used in some pattern recognition applications [3]. The FMT is generally implemented by performing a log-polar mapping followed by the Fourier transform (FT) [3]. The main idea behind this approach is to represent rotation and scaling as translations along some axes and to take advantage of the translation invariance property of the Fourier Transform.

Ordinary 2D cepstrum of a 2D signal is defined as the inverse Fourier Transform of the logarithmic spectrum of the signal and it is computed using 2D FFT. As a result it is independent of pixel amplitude variations or gray-scale changes, which leads to robustness against illumination variations. Since it is a FT based

* This work is supported by European Commission Seventh Framework Program with EU Grant: 244088(FIRESENSE)

method it is also independent of translational shifts [3]. 2D mel-cepstrum which is based on logarithmic decomposition of frequency domain grid also has the same shift and amplitude invariance properties as the 2D cepstrum. In addition, 2D Mellin-cepstrum has rotation and amplitude invariance properties. The proposed feature extraction technique is applied to the face recognition problem that is still an active and popular area of research. It should be pointed out that our aim is not the development of a complete face recognition system but to illustrate the advantages of the 2-D cepstral domain features.

In 2D mel-cepstrum and Mellin-cepstrum, the logarithmic division of the 2D DFT grid provides the dimensionality reduction. This is also an intuitively valid representation as most natural images are low-pass in nature. Unlike the ordinary Fourier or DCT domain features high-frequency DFT and DCT coefficients are not discarded in an ad-hoc manner. They are combined in bins of frequency values in a logarithmic manner during the 2D mel-cepstrum computation.

The rest of the paper is organized as follows. In Section 2, proposed 2D cepstral domain feature extraction methods are described. In Section 3, the well-known classification method SVM and a subspace based pattern recognition method called Common Matrix Approach are briefly explained. In Section 4, experimental results are presented.

2 The 2D Mel- and Mellin-Cepstrum

In the literature, the 2D cepstrum was used for shadow detection, echo removal, automatic intensity control, enhancement of repetitive features and cepstral filtering [1, 2]. In this article, 2D mel-cepstrum and Mellin-cepstrum are used for representing images or image regions.

We first introduce the 2-D mel-cepstrum using the definition of 2D cepstrum which is defined as follows. 2D cepstrum $\hat{y}(p, q)$ of a 2D image $y(n_1, n_2)$ is given by

$$\hat{y}(p, q) = F_2^{-1}(\log(|Y(u, v)|^2)) \quad (1)$$

where (p, q) denotes 2D cepstral quefrequency coordinates, F_2^{-1} denotes 2D Inverse Discrete-Time Fourier Transform (IDTFT) and $Y(u, v)$ is the 2D Discrete-Time Fourier Transform (DTFT) of the image $y(n_1, n_2)$. In practice, Fast Fourier Transform (FFT) algorithm is used to compute DTFT.

In 2D mel-cepstrum the DTFT domain data is divided into non-uniform bins in a logarithmic manner and the energy $|G(m, n)|^2$ of each bin is computed as follows

$$|G(m, n)|^2 = \sum_{k, l \in B(m, n)} |Y(k, l)|^2 \quad (2)$$

where $Y(k, l)$ is the Discrete Fourier Transform (DFT) of $y(n_1, n_2)$, and $B(m, n)$ is the (m, n) - th cell of the logarithmic grid. Cell or bin sizes are smaller at low frequencies compared to high-frequencies. This approach is similar to the mel-cepstrum computation in speech processing. Similar to speech signals most natural images including face images are low-pass in nature. Therefore, there is

more signal energy at low-frequencies compared to high frequencies. Logarithmic division of the DFT grid emphasizes high frequencies. After this step 2D mel-frequency cepstral coefficients $\hat{y}_m(p, q)$ are computed using either inverse DFT or DCT as follows

$$\hat{y}_m(p, q) = F_2^{-1}(\log(|G(m, n)|^2)) \quad (3)$$

The size of the Inverse DFT (IDFT) is smaller than the size of the forward DFT used to compute $Y(k, l)$ because of the logarithmic grid. It is also possible to apply different weights to different bins to emphasize certain bands as in speech processing. Since several DFT values are grouped together in each cell, the resulting 2D mel-cepstrum sequence computed using the IDFT has smaller dimensions than the original image. Steps of the 2D mel-cepstrum based feature extraction scheme is summarized below.

- N by N 2D DFT of input images are calculated. The DFT size N should be larger than the image size. It is better to select $N = 2^r > \text{dimension}(y(n_1, n_2))$ to take advantage of the FFT algorithm during DFT computation.
- The non-uniform DTFT grid is applied to the resultant DFT matrix and the energy $|G(m, n)|^2$ of each cell is computed. Each cell of the grid can be also weighted with a coefficient. The new data size is M by M where $M \leq N$. In most images, edges and important facial features generally contribute to high frequencies. In order to extract better representative features, high frequency component cells of the 2D DFT grid is multiplied with higher weights compared to low frequency component bins in the grid. As a result, high frequency components are further emphasized.
- Logarithm of cell energies $|G(m, n)|^2$ are computed.
- 2D IDFT or 2D IDCT of the M by M data is computed to get the M by M mel-cepstrum sequence.

It is possible to achieve illumination invariance in cepstral domain because of the logarithm operation during cepstrum computation.

Fourier-Mellin features are rotation, scale and translation invariant [3]. The 2D Mellin cepstrum feature extraction technique is a modified version of the 2D mel-cepstrum algorithm. It takes advantage of the Mellin transform and provides rotation, scale and illumination invariant features. Steps of 2D-Mellin cepstrum computation is summarized below:

- N by N 2D DFT of input images are calculated. The DFT size N should be larger than the image size. It is better to select $N = 2^r > \text{dimension}(y(n_1, n_2))$ to take advantage of the FFT algorithm during DFT computation.
- Logarithm of magnitudes of the DFT coefficients are computed.
- Non-uniform DFT grid is applied to the resultant matrix and the mean of each cell is computed. Each cell of the grid is represented with this mean and the cell is weighted with a coefficient. The new data size is M by M where $M \leq N$.
- Cartesian to Log-polar conversion is performed using bilinear interpolation. This is the key step of the Fourier-Mellin transform providing rotation and scale invariance.

- 2D IDFT of the M by M log-polar data is computed.
- Absolute value or the magnitude of the IDFT coefficients are calculated to get the M by M Mellin-cepstrum sequence.

Invariance of cepstrum to the pixel amplitude changes is an important feature. In this way, it is possible to achieve robustness to illumination invariance. Let $Y(u, v)$ denote the 2D DTFT of a given image matrix $y(n_1, n_2)$ and $cy(n_1, n_2)$ has a DTFT $cY(u, V)$ for any real constant c . The log spectrum of $cY(u, V)$ is given as follows

$$\log(|cY(u, v)|) = \log(|c|) + \log(|Y(u, v)|) \quad (4)$$

and the corresponding cepstrum is given as follows

$$\psi(p, q) = \hat{a}\delta(p, q) + \hat{y}(p, q) \quad (5)$$

where $\delta(p, q) = 1$ for $p = q = 0$ and $\delta(p, q) = 0$ otherwise. Therefore, the cepstrum values except at $(0, 0)$ location (DC Term) do not vary with the amplitude changes. Since the Fourier Transform magnitudes of $y(n_1, n_2)$ and $y(n_1 - k_1, n_2 - k_2)$ are the same, the 2D cepstrum and mel-cepstrum are shift invariant features.

Another important characteristic of 2D cepstrum is symmetry with respect to $\hat{y}[n_1, n_2] = \hat{y}[-n_1, -n_2]$. As a result only a half of the 2-D cepstrum or MxM 2-D mel-cepstrum coefficients are enough when IDFT is used.

3 Feature Classification

In this article, Common Matrix Approach (CMA) and multi-class SVM are used in feature classification. The CMA directly uses feature matrices as input. The Common Matrix Approach (CMA) is a 2D extension of Common Vector Approach (CVA), which is a subspace based pattern recognition method [4]. In this article, the CMA method is implemented as given in the reference [5] and used as a classification engine. On the other hand SVM needs a matrix to vector conversion process to convert the 2D cepstral domain feature matrices to vectors. SVM is a supervised machine learning method based on the statistical learning theory. The method constructs a hyperplane or a set of hyperplanes in a high dimensional space that can be used in classification tasks. In this work, SVM with a multi class classification support [6] with RBF kernel is used. The multi-class classification method uses “one-against-one” strategy [6].

4 Experimental Results

In this paper, AR Face Image Database [7], ORL Face Database [8] and Yale Face Database [9] are used to demonstrate the effectiveness of the proposed features. AR face database created by Aleix Martinez and Robert Benavente contains 4000 facial images of 126 subjects. In this work, 14 non-occluded poses of 50

subjects are used. The second database used in this work is ORL face database. The ORL database contains 40 subject and each subject has 10 poses. In this article 9 poses of each subject are used. The last database used in this work is the Yale Face Database. The database contains 165 facial images belonging to 15 subjects.

In order to compare performances of various features, the proposed 2D cepstral domain features, 2D Fourier Mellin Transform (FMT) based features, actual image pixel matrices, and 2D PCA based features are applied to CMA and multi-class SVM as inputs. In order to achieve robustness in recognition results, “leave-one-out” procedure is used.

In the calculation of 2D cepstral domain features, different non-uniform grids are used. Due to these different non-uniform grids, new M by M 2D cepstrum based features are generated. ($M = 49, 39, 35, 29$). The 2D cepstrum based features giving the best performance are used in the comparison with the FMT based features, 2D PCA features and actual image matrices. The size of the cepstral features given in Table 1 and Table 2 differ for that purpose.

Actual image pixel matrices, 2D PCA based feature matrices, 2D FMT and 2D cepstrum based feature matrices are applied to the CMA. These features are also applied to SVM by converting these feature matrices to feature vectors. For each face database, average recognition rates of both classifiers are obtained and displayed in the Table 1 and Table 2.

Table 1. Recognition Rates (RR) of CMA classifier with different feature sets.

Features	Face Databases					
	AR		ORL		YALE	
	RR	Feature Size	RR	Feature Size	RR	Feature Size
Original Images	97.42%	100×85	98.33%	112×92	71.52%	152×126
2D PCA	97.71%	100×12	98.33%	112×15	71.52%	152×9
2D FMT	98.28%	60×60	98.61%	60×60	73.33%	60×60
Proposed 2D Mel-Cepstrum	99%	20×39	99.44%	18×35	77.57%	20×39
Proposed 2D Mellin Cepstrum	99.28%	25×49	100%	15×29	77.57%	20×39

Based on the above experiments, the 2D PCA and 2D FMT based features do not provide better results than the proposed cepstrum based features. The Yale Face database contains face images having large illumination variations. Since CMA can not cope with large illumination changes, the recognition rates become significantly lower than the rates obtained by using SVM.

The computational complexity of 2D PCA and 2D FMT based features are higher than 2D mel-cepstrum based features which are computed using FFT. The cost of computing a 2D mel-cepstrum sequence for an N by N image is $O(N^2 \log(N)) + M^2 \log(M)$ and an additional $M^2/2$ logarithm computations which can be implemented using a look-up table. 2D Mellin-cepstrum requires an additional log-polar conversion step in the Fourier domain.

Table 2. Recognition Rates (RR) of SVM based classifier with different feature sets.

Features	Face Databases					
	AR		ORL		YALE	
	RR	Feature Size	RR	Feature Size	RR	Feature size
Original Images	96.85%	8500	98.05%	10304	88.00%	19152
2D PCA	96.85%	1200	98.33%	1680	87.87%	1368
2D FMT	97.85%	3600	98.61%	3600	90.90%	3600
Proposed 2D mel-cepstrum	98.71%	630	98.61%	630	94.54%	780
Proposed 2D Mellin-cepstrum	98.85%	630	99.44%	435	96.96%	780

5 Conclusion

In this article, a 2D mel-cepstrum and Mellin-cepstrum based feature extraction techniques are proposed for image representation. Illumination invariance and invariance to translational shifts are important properties of 2D mel-cepstrum and 2D cepstrum. In addition, 2D Mellin-cepstrum provides robustness against rotation and scale invariance. 2D Cepstral domain features extraction techniques provide not only better recognition rates but also dimensionality reduction in feature matrix sizes in the face recognition problem. Our experimental studies indicate that 2D cepstral methods are superior to classical feature extraction baseline methods in facial image representation with lower computational complexity.

References

1. Toreyin, B.U., Cetin, A.E.: Shadow detection using 2D cepstrum. Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIII **7338** (2009) 733809
2. Lee, J.K., Kabrisky, M., Oxley, M.E., Rogers, S.K., Ruck, D.W.: The complex cepstrum applied to two-dimensional images. Pattern Recogn. **26** (1993) 1579 – 1592
3. Gueham, M., Bouridane, A., Crookes, D., Nibouche, O.: Automatic recognition of shoeprints using fourier-mellin transform. In: Adapt. Hardw. and Sys., 2008. AHS '08. NASA/ESA Conf. (2008) 487 –491
4. Gulmezoglu, M., Dzharafarov, V., Barkana, A.: The common vector approach and its relation to principal component analysis. IEEE T Speech Audi. P. **9** (2001) 655–662
5. Turhal, U.C., Gulmezoglu, M.B., Barkana, A.: Face recognition using common matrix approach. In: European Signal Processing Conference. (2005)
6. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001)
7. Martinez, A., Benavente, R.: The AR face database. CVC Tech. Report # 24 (1998)
8. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: App. Comput. Vision, 1994., Proc. Second IEEE Workshop. (2002) 138–142
9. Yale: YaleFace Database. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html> (1997)